

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
7 October 2004 (07.10.2004)

PCT

(10) International Publication Number
WO 2004/086721 A2

(51) International Patent Classification⁷: **H04L 29/06**

(21) International Application Number:
PCT/GB2004/001253

(22) International Filing Date: 23 March 2004 (23.03.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0306973.9 26 March 2003 (26.03.2003) GB

(71) Applicant (for all designated States except US): **BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY** [GB/GB]; 81 Newgate Street, London, Greater London EC1A 7AJ (GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **ALVAREZ**

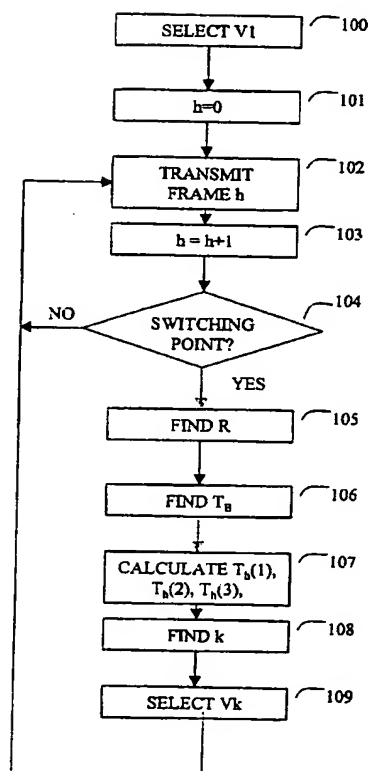
AREVALO, Roberto [ES/GB]; 4 Withipoll Street, Ipswich, Suffolk IP4 2BZ (GB). **TURNBULL, Rory, Stewart** [GB/GB]; 12 Mannington Close, Rushmere St Andrew, Ipswich, Suffolk IP4 5PW (GB). **WALKER, Matthew, David** [GB/GB]; 9 Thirlmere Court, Ipswich, Suffolk IP11 9SN (GB).

(74) Agent: **LLOYD, Barry, George, William**; BT Group Legal Intellectual Property Department, PP: C5A, BT Centre, 81 Newgate Street, London, Greater London EC1A 7AJ (GB).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM,

[Continued on next page]

(54) Title: TRANSMITTING OVER A NETWORK



(57) Abstract: Data for presentation in real time, such as a video or audio sequence, is available on different encoded versions having different degrees of compression. In order to assess, during transmission of one version, the feasibility of switching to another version, given the data rate known to be available at the time, a server computes, for a candidate version, in respect of at least one portion thereof that has not yet been sent, the maximum value of a timing error that would occur if any number of portions starting with that portion to be sent at the available rate. The selection of the same or a different version for continuing transmission is taken in dependence on a comparison between the computed error and the current state of a receiving buffer. Error values may be computed in advance for a range of transmitting rates, stored and later retrieved for use in estimating an error value corresponding to the actual transmitting rate.



TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK,

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

1
2(PRTS)10/549912
JC17 Rec'd PCT/PTO 20 SEP 2005Transmitting over a Network

The present invention is concerned with methods and apparatus for transmitting encoded video, audio or other material over a network.

- 5 According to one aspect of the present invention there is provided a method of transmitting an encoded sequence over a network to a terminal, comprising: storing a plurality of encoded versions of the same sequence, wherein each version comprises a plurality of discrete portions of data and each version corresponds to a respective different degree of compression; transmitting a current one of said versions;
- 10 ascertaining the data rate permitted by the network; ascertaining the state of a receiving buffer at the terminal; for at least one candidate version, computing in respect of at least one discrete portion thereof as yet unsent the maximum value of a timing error that would occur were any number of portions starting with that portion to be sent at the currently ascertained permitted rate; comparing the determined maximum error
- 15 values with the ascertained buffer state; selecting one of said versions for transmission, in dependence on the results of said comparisons; and transmitting the selected version.

- In another aspect, the invention provides a method of transmitting an encoded sequence over a network to a terminal, comprising: storing a plurality of encoded
- 20 versions of the same sequence, wherein each version comprises a plurality of discrete portions of data and each version corresponds to a respective different degree of compression; for each version and for each of a plurality of nominal transmitting rates, computing in respect of at least one discrete portion thereof the maximum value of a timing error that would occur were any number of portions starting with that portion to
- 25 be sent at the respective nominal rate; storing said maximum error values; transmitting a current one of said versions; ascertaining the data rate permitted by the network; ascertaining the state of a receiving buffer at the terminal; for at least one candidate version, using the ascertained permitted data rate and the stored maximum error values to estimate a respective maximum error value corresponding to said
- 30 ascertained permitted data rate; comparing the estimated maximum error values with the ascertained buffer state; selecting one of said versions for transmission, in dependence on the results of said comparisons; and transmitting the selected version.

Further aspects of the invention are set out in the claims

Some embodiments of the invention will now be described, by way of example, with reference to the accompanying drawings, in which:

Figure 1 is a block diagram of a transmission system embodying the invention;

Figure 2 is a timing diagram; and

5 Figure 3 is a flowchart explaining the operation of the control unit shown in Figure 1.

In Figure 1, a streamer 1 contains (or has access to) a store 11 in which are stored files each being a compressed version of a video sequence, encoded using a conventional compression algorithm such as that defined in the ITU standard H.261 or H.263, or one of the ISO MPEG standards. More particularly, the store 11 contains, for the same
10 original video material, several files each encoded with a different degree of compression. In practice all the material could if desired be stored in one single file, but for the purposes of description they will be assumed to be separate files. Thus Figure 1 shows three such files: V1, encoded with a high degree of compression and hence low bit-rate, representing a low-quality recording; V2, encoded with a lesser
15 degree of compression and hence higher bit-rate, representing a medium-quality recording; and V3, encoded with a low degree of compression and hence even higher bit-rate, representing a high-quality recording. Naturally one may store similar multiple recordings of further video sequences, but this is not important to the principles of operation.

20 By "bit-rate" here is meant the bit-rate generated by the original encoder and consumed by the ultimate decoder; in general this is not the same as the rate at which the streamer actually transmits, which will be referred to as the transmitting bit-rate. It should also be noted that these files are generated at a variable bit-rate (VBR) - that is, the number of bits generated for any particular frame of the video depends on the
25 picture content. Consequently, references above to low (etc.) bit-rate refer to the average bit-rate.

The server has a transmitter 12 which serves to output data via a network 2 to a terminal 3. The transmitter is conventional, perhaps operating with a well known protocol such as TCP/IP. A control unit 13 serves in conventional manner to receive
30 requests from the terminal for delivery of a particular sequence, and to read packets of data from the store 11 for sending to the transmitter 12 as and when the transmitter is able to receive them. Here it is assumed that the data are read out as discrete packets, often one packet per frame of video, though the possibility of generating more

than one packet for a single frame is not excluded. (Whilst is in principle possible for a single packet to contain data for more than one frame, this is not usually of much interest in practice).

5 Note that these packets are not necessarily related to any packet structure used on the network 2.

The terminal 3 has a receiver 31, a buffer 32, primarily for accommodating short-term fluctuations in network delay and throughput, and a decoder 33. In principle, the terminal is conventional, though to get full benefit from the use of the server, one might choose to use a terminal having a larger buffer 32 than is usual.

10 Some networks (including TCP/IP networks) have the characteristic that the available transmitting data rate fluctuates according to the degree of loading on the network. The reason for providing alternative versions V1, V2, V3 of one and the same video sequence is that one may choose a version that the network is currently able to support. Another function of the control unit 13, therefore, is to interrogate the
15 transmitter 12 to ascertain the transmitting data rate that is currently available, and take a decision as to which version to send. Here, as in many such systems, this is a dynamic process: during the course of a transmission the available rate is continually monitored so that as conditions improve (or deteriorate) the server may switch to a higher (or lower) quality version. Sometimes (as in TCP/IP) the available transmitting
20 rate is not known until after transmission has begun; one solution is always to begin by sending the lowest-rate version and switch up if and when it becomes apparent that a higher quality version can be accommodated.

Some systems employ additional versions of the video sequence representing transitional data which can be transmitted between the cessation of one version and
25 the commencement of a different one, so as to bridge any incompatibility between the two versions. If required, this may be implemented, for example, in the manner described in our U.S. patent 6,002,440.

In this description we will concentrate on the actual decision on if and when to switch. Conventional systems compare the available transmitting bit-rate with the *average* bit-
30 rates of the versions available for transmission. We have recognised, however, that this is unsatisfactory for VBR systems because it leaves open the possibility that at some time in the future the available transmitting bit-rate will be insufficient to accommodate short-term fluctuations in instantaneous bit-rate as the latter varies with picture content. Some theoretical discussion is in order at this point.

As shown in Figure 2, an encoded video sequence consists of N packets. Each packet has a header containing a time index t_i ($i=0 \dots N-1$) (in terms of real display time – e.g. this could be the video frame number) and contains b_i bits. This analysis assumes that packet i must be completely received before it can be decoded (i.e. one must buffer the whole packet first).

In a simple case, each packet corresponds to one frame, and the time-stamps t_i increase monotonically, that is, $t_{i+1} > t_i$ for all i . If however a frame can give rise to two or more packets (each with the same t_i) then $t_{i+1} \geq t_i$. If frames can run out of capture-and-display sequence (as in MPEG) then the t_i do not increase monotonically. Also, in practice, some frames may be dropped, so that there will be no frame for a particular value of t_i .

These times are relative. Suppose the receiver has received packet 0 and starts decoding packet 0 at time $t_{ref} + t_0$. At "time now" of $t_{ref} + t_g$ the receiver has received packet t_g (and possibly more packets too) and has just started to decode packet g .

Packets g to $h-1$ are in the buffer. Note that (in the simple case) if $h = g + 1$ then the buffer contains packet g only. At time $t_{ref} + t_j$ the decoder is required to start decoding packet j . Therefore, at that time $t_{ref} + t_j$ the decoder will need to have received all packets up to and including packet j .

The time available from now up to $t_{ref} + t_j$ is $(t_{ref} + t_j) - (t_{ref} + t_g) = t_j - t_g$. (1)

The data to be sent in that time are that for packets h to j , viz.

$$\sum_{i=h}^j b_i \quad (2)$$

which at a transmitting rate R will require a transmission duration

$$\frac{\sum_{i=h}^j b_i}{R} \quad (3)$$

This is possible only if this transmission duration is less than or equal to the time available, i.e. when the currently available transmitting rate R satisfies the inequality

$$\sum_{i=h}^j \frac{b_i}{R} \leq t_j - t_g \quad (4)$$

Note that this is the condition for satisfactory reception and decoding of frame j : satisfactory transmission of the whole of the remaining sequence requires that this condition be satisfied for all $j = h \dots N-1$.

For reasons that will become apparent, we rewrite Equation (4) as:

$$\sum_{i=h}^j \frac{b_i}{R} - (t_j - t_{h-1}) \leq t_{h-1} - t_g \quad (5)$$

Note that $t_j - t_{h-1} = \sum_{i=h}^j (t_i - t_{i-1}) = \sum_{i=h}^j \Delta t_i$ where $\Delta t_i = t_i - t_{i-1}$.

Also, we define $\Delta \varepsilon_i = (b_i / R) - \Delta t_i$

and $T_B = t_{h-1} - t_g$; note that T_B is the difference between the time-stamp of the most recently received packet in the buffer and the time stamp of the least recently received packet in the buffer – i.e. the one that we have just started to decode. Thus, T_B indicates the amount of buffered information that the client has at time t_g .

Then the condition is

$$\sum_{i=h}^j \Delta \varepsilon_i \leq T_B \quad (6)$$

For a successful transmission up to the last packet $N-1$, this condition must be satisfied for any possible j , viz.

$$\text{Max}_{j=h}^{j=N-1} \left\{ \sum_{i=h}^j \Delta \varepsilon_i \right\} \leq T_B \quad (7)$$

The left-hand side of Equation (7) represents the maximum timing error that may occur from the transmission of packet h up to the end of the sequence, and the condition states, in effect that this error must not exceed the ability of the receiver buffer to accommodate it, given its current contents. For convenience, we will label the left-hand side of Equation (7) as T_h - i.e.

$$T_h = \text{Max}_{j=h}^{j=N-1} \left\{ \sum_{i=h}^j \Delta \varepsilon_i \right\} \quad (8)$$

So that Equation (7) may be written as

$$T_h \leq T_B \quad (9)$$

In practice we prefer to allow switching only at certain defined "switching points" in the sequence (and naturally provide the transitional data mentioned earlier only for such points). In that case the test needs to be performed only at such points.

Figure 3 is a flowchart showing operation of the control unit 13 following selection of a video sequence for transmission. At step 100, a version, such as V1, is selected for transmission. The currently selected version number is stored. At step 101 a frame counter is reset. Then (102) the first frame (or on subsequent iterations, the next frame) of the currently selected version, is read from the store 11 and sent to the transmitter 12. Normally, the frame counter is incremented at 103 and control returns to step 102 where, as soon as the transmitter is ready to accept it, a further frame is read out and transmitted. If, however the frame is designated as a switching frame the fact that it contains a flag indicating this is recognised at step 104.

The switching decision at frame i may then proceed as follows:

Step 105: interrogate the transmitter 12 to determine the available transmitting rate R ;
 Step 106: ascertain the current value of T_B : this may be calculated at the terminal and transmitted to the server, or may be calculated at the server (see below);
 Step 107: compute (for each file V1, V2, V3) T_h in accordance with Equation (8) - let these be called $T_h(1)$, $T_h(2)$, $T_h(3)$;
 Step 108: determine the highest value of k for which $T_h(k) + \Delta \leq T_B$, where Δ is a fixed safety margin;
 Step 109: select file V_k for transmission.

The original loop is then resumed with step 102 where the next frame is transmitted before, but possibly from a different one of the three files V1, V2, V3.

The calculation of T_B at the server will depend on the exact method of streaming that is in use. Our preferred method is (as described our in international patent application no. PCT/GB 01/05246 [Agent's Ref. A26079]) to send, initially, video at the lowest quality, so that the terminal may immediately start decoding whilst at the same time the receiving buffer can be filling up because data is being sent at a higher rate than it is

used. In this case the server can deduce current client session time (i.e. the timestamp of the packet currently being decoded at the terminal) without any feedback, and so

T_B = latest sent packet time – current client session time.

If the system is arranged such that the terminal waits until some desired state of buffer fullness is reached before playing begins, then the situation is not quite so simple because there is an additional delay to take into account. If this delay is fixed, it can be included in the calculation. Similarly, if the terminal calculates when to start playing and both the algorithm used, and the parameters used by the algorithm, are known by the server, again this can be taken into account. If however the terminal is of unknown type, or controls its buffer on the basis of local conditions, feedback from the terminal will be needed.

Now, this procedure will work perfectly well, but does involve a considerable amount of processing that has to be carried out during the transmission process. In a modified implementation, therefore, we prefer to perform as much as possible of this computation in advance. In principle this involves the calculation of $T_h(k)$ for every packet that follows a switching point, and storing this value in the packet header. Unfortunately, this calculation (Equation (8) and the definition of $\Delta \epsilon_i$) involves the value of R , which is of course unknown at the time of this pre-processing. Therefore we proceed by calculating $T_h(k)$ for a selection of possible values of R , for example (if R_A is the average bit rate of the file in question)

$$R_1 = 0.5R_A$$

$$R_2 = 0.7R_A$$

$$R_3 = R_A$$

$$R_4 = 1.3R_A$$

$$R_5 = 2R_A$$

So each packet h has these five precalculated values of T_h stored in it. If required (for the purposes to be discussed below) one may also store the relative time position at which the maximum in Equation (8)) occurs, that is,

$\Delta t_{h \max} = t_{j \max} - t_h$ where $t_{j \max}$ is the value of j in Equation 8 for which T_h is obtained.

In this case the switching decision at frame h proceeds as follows:

interrogate the transmitter 12 to determine the available transmitting rate R ;

ascertain the current value of T_B , as before;

EITHER - in the event that R corresponds to one of the rates for which T_h has been precalculated - read this value from the store (for each file V1, V2, V3);

OR - in the event that R does not so correspond, read from the store the value of T_h (and, if required, $t_{h\max}$) that correspond to the highest one (R^-) of the rates $R_1 \dots R_J$ that is

5 less than the actual value of R , and estimate T_h from it (again, for each file V1, V2, V3);
determine the highest value of k for which $T_h(k) + \Delta \leq T_B$, where Δ is a fixed safety margin;

select file V_k for transmission.

10 The estimate of T_h could be performed simply by using the value T_h^- associated with R^- ; this would work, but since it would overestimate T_h it would result, at times, in a switch to a higher quality stream being judged impossible even though it were possible. Another option would be by linear (or other) interpolation between the values of T_h stored for the two values of $R_1 \dots R_J$ each side of the actual value R . However, our preferred approach is to calculate an estimate according to:

$$15 \quad T_i' = \frac{(T_i^- + \Delta T_{i\max}^-) R^-}{R} - \Delta T_{i\max}^-$$

Where R^- is the highest one of the rates $R_1 \dots R_J$ that is less than the actual value of R , T_i^- is the precalculated T_h for this rate, $\Delta T_{i\max}^-$ is the time from t_i at which T_i^- is obtained (i.e. is the accompanying value of $\Delta t_{h\max}^-$. In the event that this method returns a negative value, we set it to zero.

20 Note that this is only an estimate, as T_h is a nonlinear function of rate. However with this method T_i' is always higher than the true value and automatically provides a safety margin (so that the margin Δ shown above may be omitted).

Note that these equations are valid for the situation where the encoding process generates two or more packets (with equal t_i) for one frame, and for the situation
25 encountered in MPEG with bidirectional prediction where the frames are transmitted in the order in which they need to be decoded, rather than in order of ascending t_i .

The above description assumes that the test represented by Equation (7) is performed for all versions of the stored video. Although preferred, this is not essential. If large jumps in picture quality are not expected (for example because frequent switching
30 points are provided) then the test could be performed only for the current version and one or more versions corresponding to adjacent compression rates. For example,

when transmitting version V1, it might be considered sufficient to perform the test only for the current version V1 and for the nearest candidate version V2. Also, in the case of a server that interfaces with different networks, one might choose to test only those versions with data rate requirements that lie within the expected range of capability of the particular network in use.

5

Although the example given is for encoded video, the same method can be applied to encoded audio or indeed any other material that is to be played in real time.

Claims

1. A method of transmitting an encoded sequence over a network to a terminal,
5 comprising
storing a plurality of encoded versions of the same sequence, wherein each version
comprises a plurality of discrete portions of data and each version corresponds to a
respective different degree of compression;
transmitting a current one of said versions;
10 ascertaining the data rate permitted by the network;
ascertaining the state of a receiving buffer at the terminal;
for at least one candidate version, computing in respect of at least one discrete portion
thereof as yet unsent the maximum value of a timing error that would occur were any
number of portions starting with that portion to be sent at the currently ascertained
15 permitted rate;
comparing the determined maximum error values with the ascertained buffer state;
selecting one of said versions for transmission, in dependence on the results of said
comparisons; and
transmitting the selected version.
- 20
2. A method of transmitting an encoded sequence over a network to a terminal,
comprising
storing a plurality of encoded versions of the same sequence, wherein each version
25 comprises a plurality of discrete portions of data and each version corresponds to a
respective different degree of compression;
for each version and for each of a plurality of nominal transmitting rates, computing in
respect of at least one discrete portion thereof the maximum value of a timing error that
would occur were any number of portions starting with that portion to be sent at the
30 respective nominal rate;

- storing said maximum error values;
- transmitting a current one of said versions;
- ascertaining the data rate permitted by the network;
- ascertaining the state of a receiving buffer at the terminal;
- 5 for at least one candidate version, using the ascertained permitted data rate and the stored maximum error values to estimate a respective maximum error value corresponding to said ascertained permitted data rate;
- comparing the estimated maximum error values with the ascertained buffer state;
- selecting one of said versions for transmission, in dependence on the results of said
- 10 comparisons; and
- transmitting the selected version.
3. A method according to claim 1 or 2 in which said maximum timing error
- 15 determination is performed only for selected ones of said portions at which a version change is to be permitted.
4. A method according to claim 1, 2 or 3 in which each computed timing error
- 20 value is the difference between (a) the time needed to transmit, at the relevant rate, the portion in question and zero or more consecutive subsequent portions up to and including any particular portion, and (b) the difference between the playing instant of the respective particular portion and the playing instant of the portion preceding the portion in question.
- 25
5. A method according to any one of claims 1 to 4 in which the sequence is a video sequence.

5. A method according to any one of claims 1 to 4 in which the sequence is an audio sequence.

- 5 7. A video recording stored on a carrier, comprising
a plurality of encoded versions of the same video sequence, wherein each version
comprises a plurality of discrete portions of data and each version corresponds to a
respective different degree of compression; and
for each discrete portion of each version and for each of a plurality of nominal
10 transmitting rates, a maximum error value for that portion, being the maximum of (a)
the value of a timing error that would occur were that portion to be sent at the
respective nominal rate; and
(b) the values of a timing error that would occur were that portion and any number of
subsequent portions subsequent thereto to be sent at the respective nominal rate.

15

8. An audio recording stored on a carrier, comprising
a plurality of encoded versions of the same audio sequence, wherein each version
comprises a plurality of discrete portions of data and each version corresponds to a
respective different degree of compression; and
20 for each discrete portion of each version and for each of a plurality of nominal
transmitting rates, a maximum error value for that portion, being the maximum of (a)
the value of a timing error that would occur were that portion to be sent at the
respective nominal rate; and
(b) the values of a timing error that would occur were that portion and any number of
25 subsequent portions subsequent thereto to be sent at the respective nominal rate.

9. An apparatus for transmitting an encoded sequence over a network to a terminal, comprising

a store storing a plurality of encoded versions of the same sequence, wherein each version comprises a plurality of discrete portions of data and each version corresponds to a respective different degree of compression;

a transmitter; and

- 5 control means operable to receive data as to the data rate permitted by the network and data as to the state of a receiving buffer at the terminal and, for at least one candidate version, to compute in respect of at least one discrete portion thereof as yet unsent the maximum value of a timing error that would occur were any number of portions starting with that portion to be sent at the permitted rate, to compare the
10 determined maximum error values with the buffer state and to select one of said versions for transmission, in dependence on the results of said comparisons.

10. An apparatus for transmitting an encoded sequence over a network to a
15 terminal, comprising

a store storing a plurality of encoded versions of the same sequence, wherein each version comprises a plurality of discrete portions of data and each version corresponds to a respective different degree of compression, each version including, for each of a plurality of nominal transmitting rates, in respect of at least one discrete portion thereof,
20 the maximum value of a timing error that would occur were any number of portions starting with that portion to be sent at the respective nominal rate;

a transmitter; and

control means for receiving data as to the data rate permitted by the network and data as to the state of a receiving buffer at the terminal and, for at least one candidate
25 version, to use the permitted data rate and the stored maximum error values to estimate a respective maximum error value corresponding to said permitted data rate, to compare the estimated maximum error values with the buffer state and to select one of said versions for transmission, in dependence on the results of said comparisons.

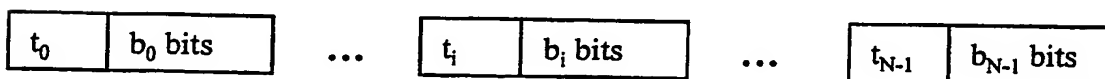
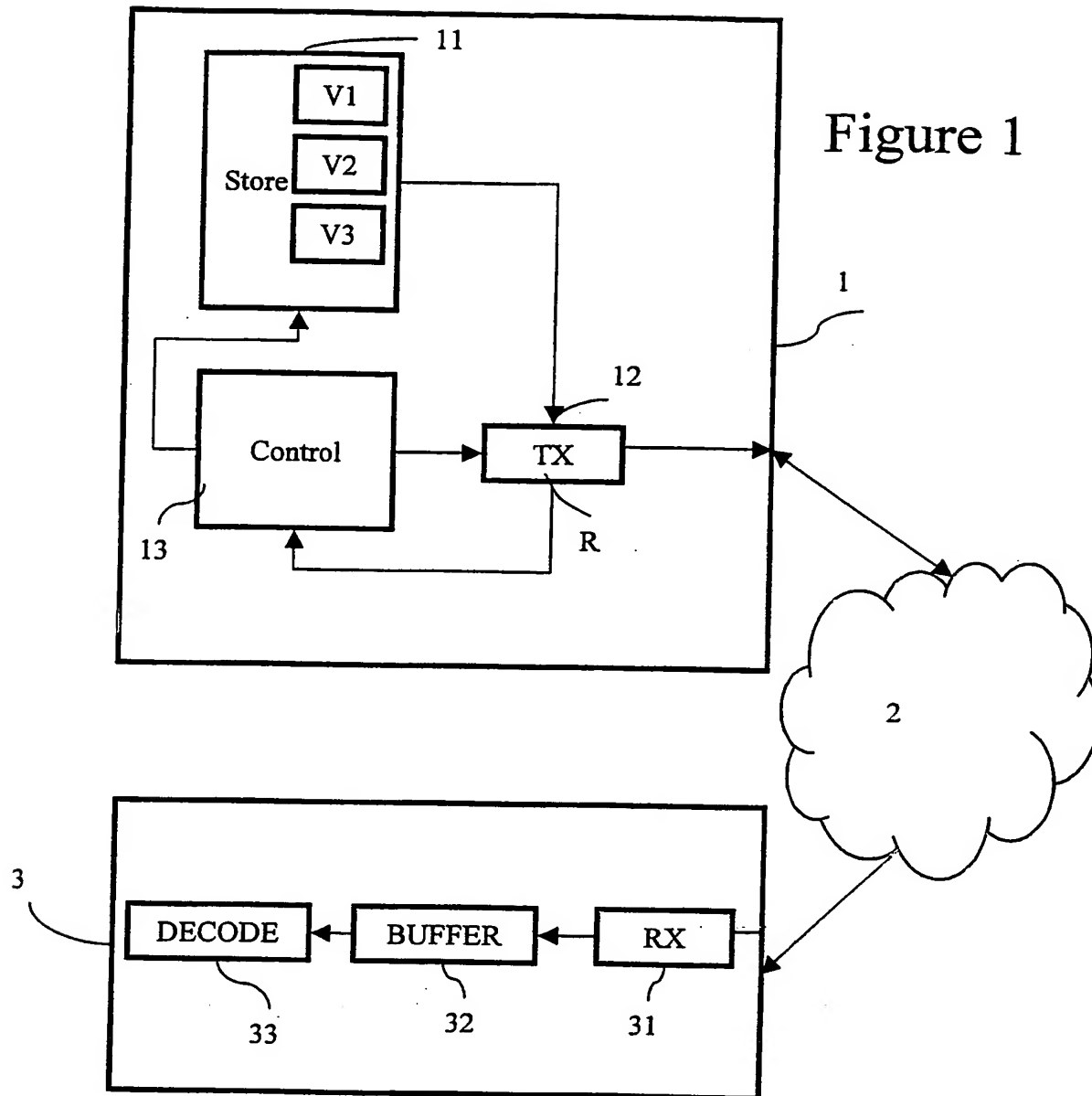


Figure 2

10/549912

2/2

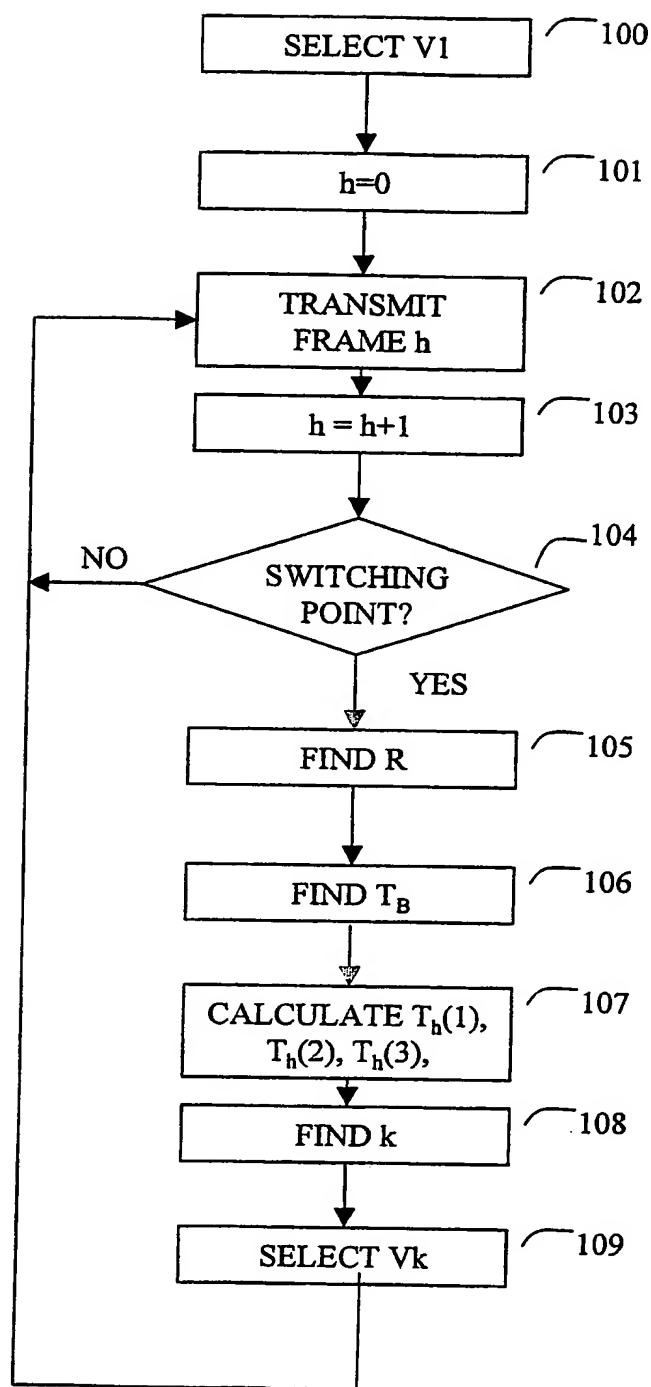


Figure 3